American Dietetic Association
Your link to nutrition and health.[sm]

120 South Riverside Plaza, Suite 2000
Chicago, IL 60606-6995
800/877-1600
www.eatright.org

Policy Initiatives and Advocacy
1120 Connecticut Avenue, Suite 480
Washington, DC 20036-3989
202/775-8277 FAX 202/775-8284

February 25, 2004

Division of Dockets Management (HFA-305)
Food and Drug Administration
Docket #: 2003N-0496
5630 Fishers Lane Rm. 1061
Rockville, MD 20852

Dear Sir or Madam:

The American Dietetic Association (ADA) represents nearly 70,000 food and nutrition professionals serving the public through the promotion of optimal nutrition, health and well being. ADA appreciates this opportunity to respond to the Food and Drug Administration's (FDA) advance notice of proposed rulemaking (ANPR) published in the November 25, 2003 *Federal Register* on alternatives for regulating qualified health claims in the labeling of conventional human foods and dietary supplements. Our comments address the specific questions posed by FDA, including how FDA should evaluate scientific evidence for qualified health claims as well as the treatment of conventional foods and dietary supplements. Included with these comments is a copy of ADA's recommendations to the Task Force on Consumer Health Information for Better Nutrition submitted in June 2003.

General Comments

ADA strongly supports the current standard of significant scientific agreement (SSA). Health and nutrient claims authorized for foods and dietary supplements should be based on the totality of publicly available scientific evidence, including results from well-designed studies conducted in a manner that is consistent with generally recognized scientific procedures and principles.

ADA commends FDA for choosing to adopt an evidence-based ranking system as the alternative to SSA for its oversight of qualified health claims. We believe such a system may help consumers make informed, science-based decisions in considering qualified health claims appearing on food labels. The food label is

an important vehicle for providing nutrition information to consumers. However, due to the complexity of nutrition, apparent conflicts in scientific approaches and outcomes, as well as the scientific community's imperfect knowledge, it is not necessarily an easy task to produce and agree upon what is truthful and nonmisleading information.

FDA may find an evidence-based approach will improve their ability to evaluate and communicate the scientific evidence behind various health claims. However, the use of evidence grading should not diminish the standard set for food, and it should be absolutely transparent for supplements. The overall effect should increase the consumer's confidence in the science being right, protecting them and guiding their decisions.

In 2001, ADA adopted an evidence-based process for evaluating different types of nutrition research for nutrition care guidelines. We attach a summary of this approach. A similar evidence-based approach has been identified for enhancing the review of the dietary guidelines [1].

ADA believes that the system we used for our Medical Nutrition Therapy Guides for Practice also has application when evaluating the strength of evidence for proposed health claims. A thorough and systematic review of the relevant science is preferable to premature approval of a claim for which limited data may exist. In addition, ADA encourages FDA to utilize the interim final rules process as long as FDA issues accompanying regulations in order to prevent any compromise of scientific integrity.

II. Health Claims

For food labeling and health claims to provide the consumer with useful information, claims should be substantiated and related to current public health problems. Health-related messages on food labels need to convey a food's relationship to a total diet over time, and they should be consistent with current dietary guidance from government entities. Both qualified and unqualified health claims should assist the public in both integrating specific food products into a well-balanced diet and avoiding unhealthful distortion of dietary habits in the hopes of preventing or curing specific chronic diseases.

Current law requires health claims and nutrient content claims to be linked with both the positive and negative attributes of food as they contribute to the total diet. Allowing foods that are high in components known to have a detrimental health effect when consumed in large quantities over time to bear any type of health claim would mislead consumers and undermine the credibility of scientific research on diet and health. For example, full fat ice cream may be high in calcium, and a health claim has been approved stating calcium consumption may

---

[1] Myers, E. Systems for evaluating nutrition research for nutrition care guidelines: Do they apply to population dietary guidelines? *J Am Diet Assoc* 2003; Suppl 2 103(12): S34

Results of consumer research conducted by the American Dietetic Association suggest consumers are increasingly aware of the link between nutrition and health, and they seek foods to reduce their risk of chronic diseases and obesity[2]. Because of this, health claims that attract consumer attention must be based on scientific evidence supporting the substance/disease relationship and not the wording of the claim.

Food and nutrition misinformation can have harmful effects on both the health and the economic status of consumers. Consumers must be provided with information they can trust is based on scientific evidence. Therefore, ADA believes Option 2 is less desirable than Option 1.

Option 3: ADA does not support

ADA does not support Option 3 because it treats qualified health claims as wholly outside of the NLEA, and it regulates them on a post-market basis. The first weakness is regulating them in a manner significantly different from unqualified health claims. It does not serve the public good to provide them with information they perceive as similar when it is regulated in a manner than poses a much stricter standard on one (unqualified health claims) than another (qualified health claims).

B. Issues Raised in the Task Force Report

Revised Claim Language for Unqualified Health Claims

FDA should thoroughly test the interpretation of the word "may" to determine how consumers understand messages they encounter when making personal food choices.

Use of Phrases Such as "FDA authorized" in Qualified and Unqualified Health Claims

ADA does not support the use of "FDA authorized" for qualified or unqualified health claims. The potential exists for consumers to be misled and confused. "FDA authorized" sounds similar to "FDA approved," which denotes safety and efficacy. When applied to dietary supplements, the statement becomes even more misleading, since unlike drugs, federal law prohibits FDA from conducting safety or efficacy tests on dietary supplements on a pre-market basis.

Consumer Education

Research by the Keystone Commission, FDA and the Commission on Dietary Supplement Labels suggests that consumers do not differentiate between labeling claims in different regulatory categories and are not likely to know that

---

[2] American Dietetic Association. *Nutrtion and You: Trends 2002.*

an unqualified health claim requires a higher substantiation and more rigorous approval process than a qualified health claim. We urge FDA to carefully consider application of an evidence-based ranking system in label claims and to develop an aggressive educational campaign to inform consumer abut the process.

Consumer education is essential to the success and effectiveness of any food package labeling system. As part of weight loss interventions, ADA members often teach patients how to read and interpret labels so they can translate label information into a continuous set of choices that will lead to a healthier diet.

We need national resources directed into consumer research to better understand how and the extent to which, labeling information is used and understood by consumers. Such research would ensure labeling information effectively promotes consumer awareness and is helpful in making purchase decisions, especially for those consumers seeking to make dietary changes to prevent or manage a chronic disease.

Evaluation of Outside Scientific Groups

ADA appreciates being cited as an organization that might be asked to evaluate scientific information and provide advice on diet and health. While resources to conduct evaluation of evidence supporting a health claim are limited, ADA is willing to consult on the process of the evidence-based review, and we can offer our assistance and recommendations for an advisory committee on the subject should FDA decide to appoint such a body of experts.

We support the current interim process being used by FDA and AHRQ to review qualified health claims. We encourage FDA to maintain the internal nature of the process, as this provides an opportunity to collaborate with other governmental agencies such as the U.S. Department of Agriculture (USDA), National Institutes of Health (NIH) and the Agency for Healthcare Research and Quality (AHRQ). By collaborating with other executive branch agencies, FDA ensures the government speaks with one voice with regard to nutrition and health issues.

III. Dietary Guidance

B. Issues Relating to Dietary Guidance

Dietary Guidance on Food Labels

Because consumers appear to rely on information on the food label when making purchases, substantive information would be useful. In considering the relationship between the food label and dietary guidance, we urge FDA to include other agencies such as USDA and the NIH Office of Dietary Supplements.

Government agencies with jurisdiction over nutrition must speak with one voice to gain the trust and attention of the population.

Conclusion

We hope these comments are useful as the agency moves forward with this initiative to facilitate and encourage the flow of high quality information on the health benefits of conventional foods and dietary supplements to consumers. Please do not hesitate to call ADA's government relations staff at (202) 775-8277 with any questions.

Sincerely,

Marianne Smith Edge/mav

Marianne Smith Edge, MS, RD, LD FADA
President

Attachments

1. American Dietetic Association. "Response to Questions Posed by the Task Force on Consumer Health Information for Better Nutrition." June 3, 2003.

2. Greer, N; Mosser, G; Logan, G and G. Halaas. "A Practical Approach to Evidence Grading." *Jt Comm J Qual Improv* 2000;26: 700-712.

Addendum
Additional References and Comments

Hancock, H.; Rogers, W.; Fisk, A. "An evaluation of warning habits and beliefs across the adult life span." *Hum Factors* 2002; 43(3): 343-354. [1]

Moore, M. "Product warning effectiveness: Perception versus Reality" *Professional Safety* 1991; 36(4): 21-24. [2]

Stutts, M; Hunnicut, G. *Journal of Advertising* 1987;16(1):41-46. [3]

Truitt, L et al. *Tob Control* 2002; June Supplement (2):59-63. [4]

The following comments from Dr. Esther Myers from the above references lend support to ADA's position articulated in this document.

1. Hancock, Rogers and Fisk reported that persons older than 55 years attended to warning symbols more often than younger consumers; however, they thought the warnings were less important.

2. Limited data is available from other sources citing lack of positive effects of warning labels in other products. The Failure Analysis Associates in Palo Alto (FAA), CA classify studies on disclaimers/warnings into two categories: qualitative and quantitative. The first would be related to surveys of users' claims of safety behavior, and the second would be based on exposure versus injury rates. While this framework was developed for a different type of product (e.g. seatbelt use and injuries related to lack of seatbelt use), it might be useful in identifying the types of empirical data that the agency should strive for. The author, summarizing the FAA article, further concludes that for a field that receives so much attention, the product warning field lacks objective data for drawing conclusions.

3. Stutts and Hunnicut define a disclaimer as a "disclosure made with the intent of clarifying potentially misleading or deceptive statements made within an advertisement."

4. Truitt et al reported that font size was an important factor in determining effectiveness of tobacco warning/disclaimers.

ICSI

INSTITUTE FOR CLINICAL SYSTEMS IMPROVEMENT

EVIDENCE GRADING SYSTEM

Development

Evidence grading was introduced into ICSI guidelines and technology assessment reports in 1996. At that time, a modification of the system used in the Agency for Health Care Policy and Research (AHCPR) Unstable Angina: Diagnosis and Management Clinical Practice Guideline was used[1]. The system called for assignment of an A (randomized, controlled trials published in peer-reviewed journals), a B (other well-designed studies published in peer-reviewed journals including cohort studies, case-control studies, trials with historic or non-randomized controls, and meta-analyses), or a C (uncontrolled case series or expert opinion) to individual research reports. No grades were assigned to other guidelines, consensus statements, or review papers. If there was "A" evidence supporting a conclusion, the conclusion grade was "A."

Based on feedback from the guideline and technology assessment work groups, it became obvious that this system was not meeting their needs. Specifically, the system was viewed as too simplistic, there were objections to grading conclusions strictly on research design type (given that quality can vary greatly within a research design type and that not all of the design types are feasible/appropriate for all research questions), there was concern that there was no consideration for how much evidence there was, and there was concern that all design types were not adequately considered.

As a result of this feedback, ICSI assembled a work group to review the evidence grading system in use at ICSI, to review other evidence grading systems available in the literature, and to make recommendations for changes to the ICSI system. The group included physicians and researchers with backgrounds in quality improvement, clinical epidemiology, and biostatistics. Recommendations formulated by the work group were submitted for approval to the ICSI committees that oversee the guideline program and the technology assessment program.

The evidence grading review work group started by establishing goals for an evidence grading system. These goals were:

1. to increase the systematic use of evidence by work groups by providing a framework and a step-by-step process for reaching key conclusions;
2. to provide a method for reaching evidence-based conclusions that busy, practicing clinicians accept as practical;

3. to provide a reliable method for grading conclusions based on the strength of the underlying evidence; and
4. to convey to readers and users of the documents the strength of the underlying evidence.

The work group reviewed many existing evidence grading systems including the system used by the United States Preventive Services Task Force[3], the system developed by Sacket[4] and modified by Cook et al.[5,6], and the system presented in the series on Users' Guides to the Medical Literature[7]. Overall, it was apparent that no one system was universally applied and that the systems varied a great deal in complexity. Although the ICSI work group decided that no one existing system fulfilled the goals identified above, there were features of the existing systems that could be incorporated into a new ICSI system. Specifically, the work group agreed that it was important to separate the evaluation of individual research reports from the assessment of the totality of evidence supporting a conclusion. The work group also agreed that assessing the quality of the individual research reports was important.

The System

The centerpiece of the evidence grading system is the conclusion grading worksheet. Conclusion grades are assigned to key conclusions and/or recommendations as determined by the guideline or technology assessment work group members. The worksheet, similar to an evidence table, is used to display and synthesize the evidence supporting a particular conclusion. An example of a worksheet from the Congestive Heart Failure guideline is presented in Figure 1. The work group formulates a tentative conclusion statement and, based on a literature search done by a medical librarian using keywords suggested by the work group, identifies the key references to include on the worksheet. The work group is encouraged to identify the strongest possible evidence (based on design type, sample size, patient population, etc.) that supports or disputes the conclusion statement. The worksheet is then prepared by ICSI staff and includes, for each reference, the citation, design type, class of research report, quality score, information about the population studied, results of the study, and the authors' conclusions. The conclusion grading worksheet is reviewed by a designated member of the work group and a tentative conclusion grade is selected. The designated work group member then presents the worksheet to the rest of the work group. There is discussion of the individual research reports and comments from the work group may be added to the worksheet. There is also discussion of the proposed conclusion grade and a final decision is made on the appropriate grade. Involvement of a member of the work group in the development of the worksheet and the deliberation by the work group in determining the final conclusion grade are considered strengths of the system. Further information about the classes of research reports, quality scores, and conclusion grades is presented in Figure 1.

Figure 1. Conclusion Grading Worksheet

Work Group's Conclusion: Digoxin improves symptoms, exercise tolerance, and quality of life, but neither increases or decreases mortality.

Conclusion Grade: I

Author/Year	Design Type	Class	Quality +,–,ø	Population Studied/Sample Size	Primary Outcome Measure(s)/Results (e.g., p-value, confidence interval, relative risk, odds ratio, likelihood ratio, number needed to treat)	Authors' Conclusions/ *Work Group's Comments (italicized)*
Captopril-Digoxin Research Group (1988)	RCT	A	ø	-Patients (<75 years) with sinus rhythm and heart failure secondary to ischemic heart disease, primary myocardial disease, or in heart failure without significant valvular regurgitation after valvular surgery (receiving diuretic therapy if needed) -Randomized to 1 of 3 groups (after withdrawal from therapy and stabilization of diuretic dose): captopril (25 mg 3x/day increased to 50 mg 3x/day after 1 wk if tolerated), digoxin (0.125-0.375 mg daily based on trough serum levels), or placebo -Included: ejection fraction ≤40%, treadmill time >4 min but < age- and sex-predicted average maximum -Excluded: MI within preceding 2 mos, unstable angina, hypertension (SBP>160 mmHg, DBP >95 mmHg) despite diuretic therapy, pulmonary disease (FEV^1/FVC ratio <60%) -Concomitant therapy with inotropic agents, vasodilators, β-adrenergic blockers, calcium antagonists, immunosuppressive agents, or other investigational drugs was prohibited	-Follow-up at 1, 2, & 6 mos after randomization -300 patients randomized (104 to captopril, 96 to digoxin, and 100 to placebo); baseline characteristics were similar (captopril group younger, p=0. 02) -Mean changes from baseline (analysis while adhering to assigned therapy): Variable — Captopril — Digoxin — Placebo Exer. time (s) — 82* n=101 — 54 n=95 — 35 n=97 NYHA class — -0.20** n=100 — -0.09 n=95 — 0.02 n=98 Eject. fraction (%) — 1.8 n=87 — 4.4*** n=82 — 0.9 n=78 Premature beats per hour# — -29.4**** n=55 — 2.3 n=47 — -16.1 n=45 *different from placebo (p<0.05); ** different from placebo with respect to proportion of patients improved (p<0.01) (see below); *** different from placebo (p<0.01) and captopril (p<0.05) groups; **** different from digoxin (p<0 05); #only patients with >10 ventricular premature beats/hr at baseline -NYHA class improved for 41% of captopril; 31% of digoxin, and 22% of placebo groups -Withdrawal from study because of treatment failure occurred with 15% of placebo group (vs. 5.8% of captopril group and 4.2% of digoxin group; p<0.05); more patients in placebo group required increase in diuretic dose (p<0.005) and hospitalization (or emergency visits) (p<0.05) than in other groups -Similar trends seen in intention-to-treat analysis -Rate of discontinuation due to adverse drug reactions: 2.9% captopril, 4.2% digoxin, 0% placebo -More possible adverse drug effects attributed to captopril (44.2%) during blinded portion of study than to other treatments (30.2% digoxin, 24% placebo) (usually mild and transient dizziness and lightheadedness) -21 deaths (8 captopril, 7 digoxin, 6 placebo)	-Captopril therapy is significantly more effective than placebo and is an effective alternative to digoxin treatment in patients with mild to moderate heart failure who are undergoing maintenance diuretic therapy. Significant improvements in exercise tolerance and functional class compared to the placebo group were seen in the captopril group but not the digoxin group. Captopril also significantly reduced ventricular premature beat rates compared with digoxin in patients with more than 10 premature beats/hour at baseline. Digoxin significantly increased left ventricular ejection fractions compared with both placebo and captopril. Patients receiving placebo had a greater incidence of treatment failure and required significantly more diuretics, hospitalizations, and/or emergency department visits for heart failure than did patients receiving captopril or digoxin. NOTES: trial was double-blind; did intention-to-treat analysis (as well as analysis while patients adhered to assigned therapy); most patients were NYHA functional class II;

Digitalis Investigation Group (1997)	RCT	A	ø	-6800 patients (302 clinical centers) with heart failure and left ventricular ejection fraction < 0.45 in normal sinus rhythm -988 patients with heart failure and ejection fraction >0.45 were enrolled in ancillary trial -May have been already receiving digoxin -Randomly assigned to digoxin or placebo (digoxin dose varied) -Other therapy used if patient had worsening symptoms of heart failure; if remained symptomatic, allowed open-label treatment with digoxin	-Follow-up visits at 4 and 16 wks then every 4 mos (mean duration 37 mos, range 28-58 mos) -No significant differences between groups (baseline) -1181 deaths in digoxin group (34.8%), 1194 in placebo group (35.1%) (RR=0.99, 95%CI: 0.91-1.07) -1016 deaths from cardiovascular causes in digoxin group (29.9%), 1004 in placebo group (29.5%) (RR=1.01; 95%CI: 0.93-1.10) -Trend toward lower risk of mortality attributable to worsening heart failure in digoxin group (p=0.06) -910 patients hospitalized for worsening heart failure in digoxin group and 1180 in placebo group (RR=0.72; 95%CI: 0.66-0.79) -Risk of death from any cause or hospitalization for worsening heart failure was lower in digoxin group (RR=0.85; 95% CI: 0.79-0.91); similar results for death due to worsening heart failure or hospitalization related to worsening heart failure -Fewer hospitalizations for any cause (per patient) in digoxin group (p=0.01) and for cardiovascular causes (p<0.001) -Benefit of digoxin appeared to be greater among patients at high risk (lower ejection fraction, enlarged heart, or NYHA III or IV) -At 1 yr, 85.6% of digoxin group patients were taking study drug and 82.9% of placebo group were taking placebo; at final study visit 70.8% of surviving patients in digoxin group were taking study drug and an additional 10.3% were taking open-label digoxin; 67.9% of surviving placebo group patients were taking placebo and 15.6% were taking open label digoxin. -Suspected digoxin toxicity greater in digoxin group -In ancillary trial, no difference in number of deaths or combined outcome of death or hospitalization due to worsening heart failure	-In patients with left ventricular ejection fractions ≤0.45 digoxin had no effect on overall mortality when added to diuretics and ACE inhibitors; the risk of hospitalization was reduced and the combined outcome of death or hospitalization attributable to worsening heart failure was also reduced. In clinical practice, digoxin therapy is likely to decrease the frequency of hospitalization but not survival. NOTES: exclusion criteria were not given in this publication (previously published); trial was double-blind; did intention-to-treat analysis; physicians were strongly encouraged to give patients ACE inhibitors; patients receiving digoxin at entry were randomly assigned with no washout period; vital status of 47 patients in digoxin group and 46 in placebo group (1.4% of total) were unknown (a sensitivity analysis assuming that either all placebo or all digoxin patient died did not change overall result)

Figure 1. This is an example of a worksheet included in the ICSI guideline on Congestive Heart Failure in Adults. The work group identifies the key research articles to be summarized on the worksheet and, once the information is entered on the worksheet, determines the appropriate conclusion statement and conclusion grade. The worksheets appear as an Appendix to the guideline.

Classes of Research Reports

Each individual research report cited in a guideline or technology assessment report is assigned a class by ICSI staff (see Table 1). Primary reports of new data collection are assigned a letter A, B, C, or D based on the design type. The hierarchy of design types (with "A" representing randomized, controlled trials etc.) is fairly consistent among evidence grading systems and reflects the fact that different study design types vary in the likelihood that an individual study will be biased[8]. Secondary reports (reports that synthesize or reflect upon collections of primary reports) are assigned an M, an R, or an X. The definitions of the various design types are those found in epidemiology textbooks[9,10,11].

Table 1. Classes of Research Reports

Primary Reports of New Data Collection	
A	randomized, controlled trial
B	cohort study
C	nonrandomized trial with concurrent or historical controls case-control study study of sensitivity and specificity of a diagnostic test population-based descriptive study
D	cross-sectional study case series case report
Reports that Synthesize or Reflect Upon Collections of Primary Reports	
M	meta-analysis systematic review decision analysis cost-benefit analysis cost-effectiveness study
R	narrative review consensus statement consensus report
X	medical opinion

Research Report Quality Categories

The quality of an individual research report is designated as plus (+), minus (-), or neutral (ø) based on the questions presented in Table 2. The quality considerations reflected in the table are considerations standardly addressed in textbooks of clinical epidemiology[10,11]. The assessment of quality is completed by ICSI staff.

Table 2. Research Report Quality Categories

PLUS (+)

Y N 1. Were the inclusion and exclusion criteria exceptionally well-defined and adhered to?

Y N 2. Were no serious questions of bias introduced in the study (e.g., through the processes of subject selection, end point selection, and observation or data collection)?

Y N 3. Does the report show a statistically significant and clinically important treatment effect or, for a negative conclusion, have high power?

Y N 4. Are the results widely generalizable to other populations?

Y N 5. Were other characteristics of a well-designed study clearly addressed in the report (e.g., treatment and control groups comparable at baseline, compliance with the intervention, use of intention to treat analysis, all important outcomes measured, statistics appropriate for study design)?

If the answer to 2 or more of the above questions is "yes", the report may be designated with a plus on the Conclusion Grading Worksheet depending on the work group's overall evaluation of the report.

MINUS (–)

Y N 1. Were the inclusion and exclusion criteria unclear or was there evidence of failure to adhere to defined criteria?

Y N 2. Were serious questions of bias introduced in the study (e.g., through the processes of subject selection, end point selection, and observation or data collection)?

Y N 3. Does the report show a statistically significant but clinically insignificant effect or, for a negative conclusion, lack power and sample size?

Y N 4. Are the results doubtfully generalizable to other populations?

Y N 5. Were other characteristics of a poorly-designed study clearly evident in the report (e.g., treatment and control groups different at baseline, low compliance with the intervention, important outcomes were not measured, inappropriate statistics for study design)?

If the answer to 2 or more of the above questions is "yes", the report may be designated with a minus symbol on the Conclusion Grading Worksheet depending on the work group's overall evaluation of the report.

NEUTRAL (Ø)

If the answers to the questions pertaining to the PLUS or MINUS criteria do not indicate that the report is exceptionally strong or exceptionally weak, the report should be designated with a neutral symbol on the Conclusion Grading Worksheet.

Conclusion Grades

Conclusions and recommendations are graded either I, II, III, or IV. Descriptions of the conclusion grades as well as examples of the types of evidence that would support a specific grade are presented in Table 3.

Table 3. Conclusion Grades

Grade I: The conclusion is supported by good evidence.

The evidence consists of results from studies of strong design for answering the question addressed. The results are both clinically important and consistent with minor exceptions at most. The results are free of any significant doubts about generalizability, bias, and flaws in research design. Studies with negative results have sufficiently large samples to have adequate statistical power.

Examples:
Supporting studies might consist of two or more randomized, controlled trials or even a single well designed, well executed trial. The evidence might also come from several smaller trials combined in a single well done meta-analysis. For a question of the soundness of a diagnostic test, the evidence might be the results of a single well done comparison of the test against an established test for the same purpose, provided that there is no evidence to the contrary. For a question of the natural history of a disease, in the absence of evidence to the contrary, the evidence might be results from a single well done prospective cohort study.

Grade II: The conclusion is supported by fair evidence.

The evidence consists of results from studies of strong design for answering the question addressed, but there is some uncertainty attached to the conclusion because of inconsistencies among the results from the studies or because of minor doubts about generalizability, bias, research design flaws, or adequacy of sample size. Alternatively, the evidence consists solely of results from weaker designs for the question addressed, but the results have been confirmed in separate studies and are consistent with minor exceptions at most.

Examples:
Supporting studies might consist of three or four randomized, controlled trials with differing results although overall the results support the conclusion. The evidence might also be the results of a single randomized, controlled trial with a clinically significant conclusion but doubtful generalizability. For a question of causation, the evidence might consist of two independent case-control studies with similar conclusions. The evidence might also consist of several careful case series reports with similar conclusions from investigators working separately.

Table 3. Conclusion Grades (continued)

Grade III: The conclusion is supported by limited evidence.

The evidence consists of results from studies of strong design for answering the question addressed, but there is substantial uncertainty attached to the conclusion because of inconsistencies among the results from different studies or because of serious doubts about generalizability, bias, research design flaws, or adequacy of sample size. Alternatively, the evidence consists solely of results from a limited number of studies of weak design for answering the question addressed.

Examples:
For a question of efficacy of medical treatment, the evidence might consist of three or four randomized trials with contradictory results or serious methodological flaws; or the evidence might consist of a single trial that used historical controls. Alternatively, for a question of efficacy, the evidence might consist of one case series report. For a question of causation, the evidence might consist of results from a single case-control study, unconfirmed by other studies.

Grade IV: The conclusion is supported only by opinion.

The support for the conclusion consists solely of the statements of informed medical commentators based on their clinical experience, unsubstantiated by the results of any research studies.

Examples:
The support might consist of a review article citing only single case reports. (If the review article cites clinical trials, cohort studies, or other stronger evidence, then that evidence should govern the assignment of the grade to the conclusion.) The support might also be an editorial, consensus report, or a position statement from a national body without citations of the results of research studies. (Again, if research studies are cited, they should govern the grade assignment.)

Summary of Process

The process for reaching a conclusion grade, specifically for a guideline in development, is summarized in Figure 2. For guidelines undergoing revision and for technology assessment reports, a similar process is followed.

Figure 2. Conclusion Grading Process for Guidelines in Development

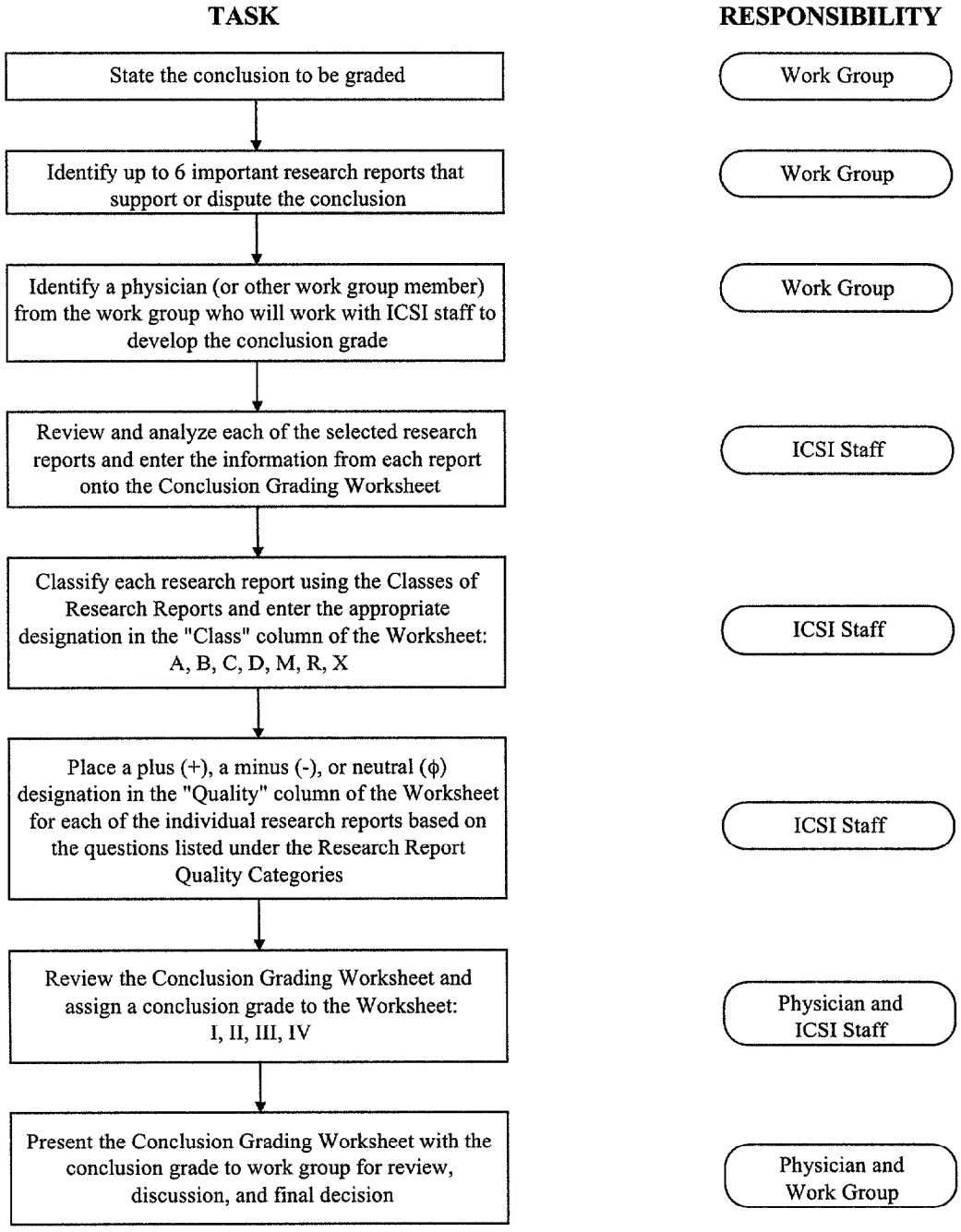

Figure 2. This figure represents the process for reaching a conclusion grade. The left column states the task to be completed and the right column identifies who is responsible for completion of that task.

Descriptions of the "Classes of Research Reports" and the "Conclusion Grades" are included in each guideline and technology assessment report. The class of research report assigned to an individual article is presented at the end of the bibliographic citation for that article. The conclusion grades are incorporated into the text of the guideline or technology assessment report with a reference to the Appendix containing the conclusion grading worksheet (see Figure 3 for an example from the ICSI Congestive Heart Failure in Adults guideline). Therefore, the reader of the document is able to use the conclusion grading information in weighing the strength of the evidence supporting the conclusion statement. This knowledge should ultimately assist the physician in making decisions about patient care.

Figure 3. Congestive Heart Failure Guideline

Discussion and References (cont)

Congestive Heart Failure in Adults

Digoxin:

Digoxin improves symptoms, exercise tolerance, and quality of life, but neither increases or decreases mortality.

Captopril-Digoxin Multicenter Research Group, The. "Comparative effects of therapy with captopril and digoxin in patients with mild to moderate heart failure." *JAMA* 259:539-44, 1988. (Class A)

Digitalis Investigation Group. "The effect of digoxin on mortality and morbidity in patients with heart failure." *N Engl J Med* 336:525-33, 1997. (Class A)

German and Austrian Xamoterol Study Group, The. "Double-blind placebo-controlled comparison of digoxin and xamoterol in chronic heart failure." *Lancet* 489-93, 1988. (Class A)

Conclusion Grade I; see Discussion Appendix D.

Figure 3. This figure presents a small portion of the Discussion and References section of the ICSI guideline on Congestive Heart Failure in Adults. The work group's conclusion statement is presented along with the references pertaining to that conclusion. The class of research report follows each reference citation. The conclusion grade and a reference to the Appendix containing the worksheet complete the section.

Guidelines and Technology Assessment reports both undergo a critical review process in which ICSI member medical groups have an opportunity to submit written critiques of the documents while still in draft form. It is expected that any critical evidence overlooked by the work group in their search of the literature would be identified during the review phase.

References

1. Greer N, Mosser G, Logan G, Halaas G. A practical approach to evidence grading. *Jt Comm J Qual Improv* 26:700-712, 2000.
2. Braunwald E, Mark DB, Jones RH, et al. Unstable angina: diagnosis and management. Clinical Practice Guideline Number 10. Agency for HealthCare Policy and Research (AHCPR), March 1994.

3. U.S. Preventive Services Task Force. Guide to clinical preventive services (2nd Edition). Baltimore: Williams and Wilkins, 1996.

4. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 95(Suppl):2S-4S, 1989.

5. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 102(Suppl):305S-311S, 1992.

6. Cook DJ, Guyatt GH, Laupacis A, Sackett DL, Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest* 104(Suppl):227S-230S, 1995.

7. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 274:1800-1804, 1995.

8. Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. *J Clin Epidemiol* 49:749-754, 1996.

9. Riegelman RK, Hirsch RP. Studying a study and testing a test. Boston: Little, Brown & Co. 1996.

10. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown & Co. 1987.

11. Friedman GD. Primer of epidemiology, 4th ed. St. Louis: McGraw-Hill, Inc. 1994.

reduce the risk of osteoporosis. However, would FDA be willing to allow a health claim saying such on ice cream, given that it is both high in saturated fat and cholesterol? The original intent of Nutrition Labeling and Education Act of 1990 (NLEA) is to protect consumers from such misleading health claims. Qualified health claims, based on less than SSA, should meet that same standard.

A. Regulatory Alternatives for Qualified Health Claims

Option 1: ADA supports

As previously stated, ADA supports strongly the SSA standard and believes all claims made on conventional foods and dietary supplements should meet this standard. The standard for food should not be negotiable because the law is clear. However, given the *Pearson v. Shalala* ruling and the task given to FDA to accommodate qualified health claims, an evidence-based grading system is a good framework given the disparate authorities for food and supplements. Of the three options, ADA most strongly supports Option 1, which would incorporate the interim procedures and evidence-based ranking system into a regulation under notice-and-comment rulemaking.

The pre-market clearance system of Option 1 provides FDA the opportunity to review qualified health claims and the supporting data, it allows the public to provide comment, and it is consistent with the intent of the Nutrition Labeling and Education Act of 1990 (NLEA). An evidence-based ranking system facilitates a thorough review of scientific data that supports both qualified and unqualified health claims.

Option 1 gives FDA the ability to revise and/or revoke a qualified health claim more quickly if new scientific evidence becomes available that alters or refutes the meaning or accuracy of any given claim. Because science is an evolutionary process, conclusions drawn at any given point in time may change as additional studies add to the pre-existing body of knowledge. Therefore, any system that relies upon an evolving scientific knowledge base needs to be flexible.

Option 2: ADA does not support

The current NLEA statute, under which authorized health claims describe a relationship between a food substance and a disease or other health-related condition, allows a health claim only if SSA among qualified experts exists about the validity of the relationship described in the claim. To reinterpret the meaning of SSA to apply to the words used in the claim itself and not the relationship undermines the intent of NLEA because we believe it is the substance/disease relationship that is meaningful to the consumer and the reason behind the purchase of the food item.